

Towards a better protein structures similarity measure

EPI Symbiose, INRIA Rennes

Rumen Andonov (randonov@irisa.fr)

Description

This project deals with the fundamental problem of comparing two three-dimensional (3D) objects. The question arises frequently and can appear in various situations and in very different domains. In bioinformatics this corresponds to computing the similarity between two protein structures and is a task of crucial importance in the domain. Proteins adopt complex 3D shapes, in which their constructive elements (amino-acids) come into proximity in 3D space and form contacts. This proximity relation is captured by a contact map -- a simple graph with vertices corresponding to the amino-acids and edges modeling the Euclidian distance between the vertices.

The approach, called contact-map-overlap (CMO), is currently extensively investigated. Computer scientists are attracted by the observation that the approach naturally leads to fundamental problems in graph theory and combinatorial optimization like Maximum Common (Edge) Subgraph or Maximum Clique problem. All these problems are NP-complete and no fast solution to them is known.

This topic is active research area in the SYMBIOSE/IRISA team as well as in the Algorithmic Computational biology group, CWI, Amsterdam. Both centers have recently developed tools for detecting protein structure similarity using the CMO approach. The researchers from both teams are now investigating to possibilities for improving these tools based on their complementarily. During this Master's project you will be involved in this challenging process. We expect candidates with strong interests in designing algorithms and in combinatorial optimization. No prior knowledge in bioinformatics is expected, but willingness to work in an interdisciplinary field. The project will be jointly supervised by the University of Rennes and by the CWI Life Sciences group, Amsterdam, Netherlands (http://www.cwi.nl/en/Algorithmic_computational_biology). In the context of this collaboration the student will share his training period between both institutions.

Bibliographie

Andonov, R., Yanev, N., and Malod-Dognin, N., An efficient lagrangian relaxation for the contact map overlap problem. In WABI '08: Proceedings of the 8th international workshop on Algorithms in Bioinformatics, pp. 162-173. Springer-Verlag, 2008.

I. Wohlers, L. Petzold, F. S. Domingues, and G. W. Klau. Aligning protein structures using distance matrices and combinatorial optimization. In Proc. of German Conference on Bioinformatics, Lecture Notes in Informatics, 2009

Caprara, B. Carr, S. Istrail, G. Lancia, B. Walenz. 1001 optimal PDB structure alignments: Integer programming methods for finding the maximum contact map overlap. Journal of Computational Biology (2004) vol. 11 (1) pp. 27-52